Z-RAM® Ultra-Dense Memory for 90nm and Below

Hot Chips 2006

David E. Fisch, Anant Singh, Greg Popov
Innovative Silicon Inc.

Outline

- Device Overview
- Operation
- Architecture Features
- Challenges
- Z-RAM Performance Factors
  - Memory Density vs. Speed
  - Pipeline Operation
  - Routing Impact
  - Active Power
  - Standby Power

Z-RAM Device Overview

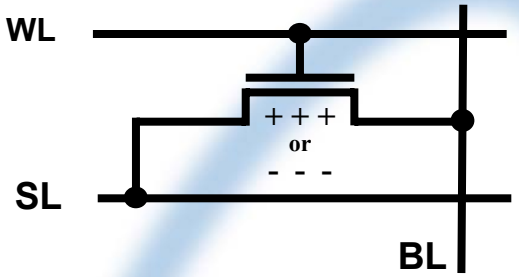
Innovative Silicon inc

- Compatible with standard SOI logic or memory process
- Small Cell Size
- As scalable as CMOS
- No new materials
- Works for N and PMOS devices
- Good SER Performance (10x better than SRAM at 65nm)
  - Small cell Size
  - SOI Substrate
  - Performance Improves with Frequency (unlike eDRAM)
- Can be extended to FD SOI, double gate etc…

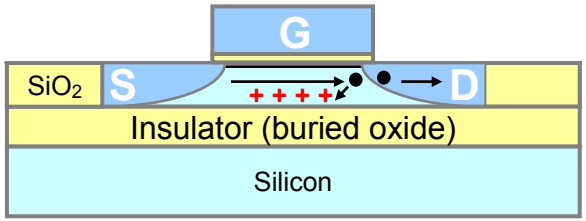
Z-RAM Device Overview: One transistor, Zero Capacitors

Innovative Silicon inc

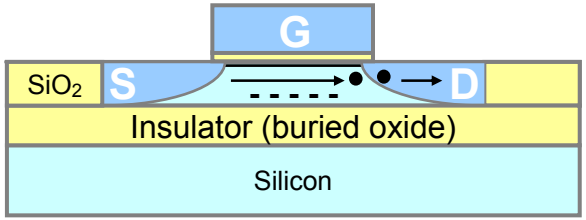
Exploits the floating body and gain effects of SOI devices

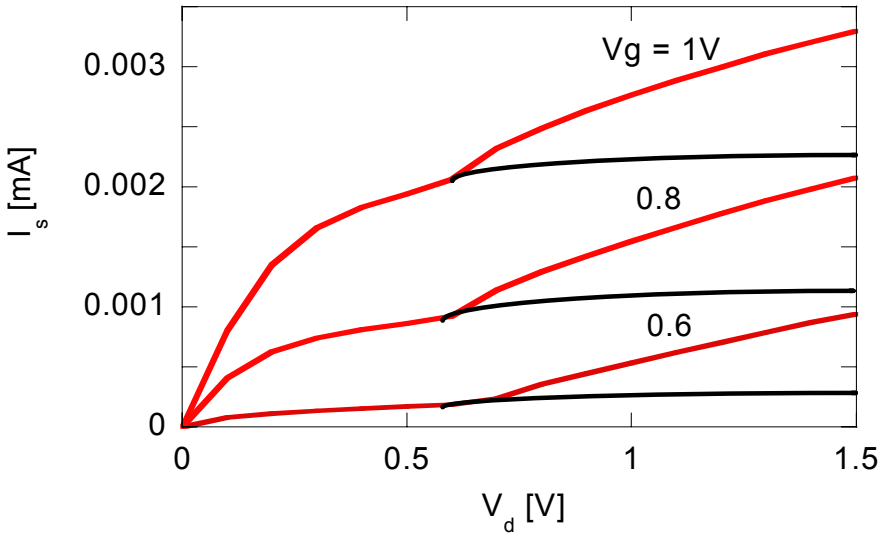


Logic "1"

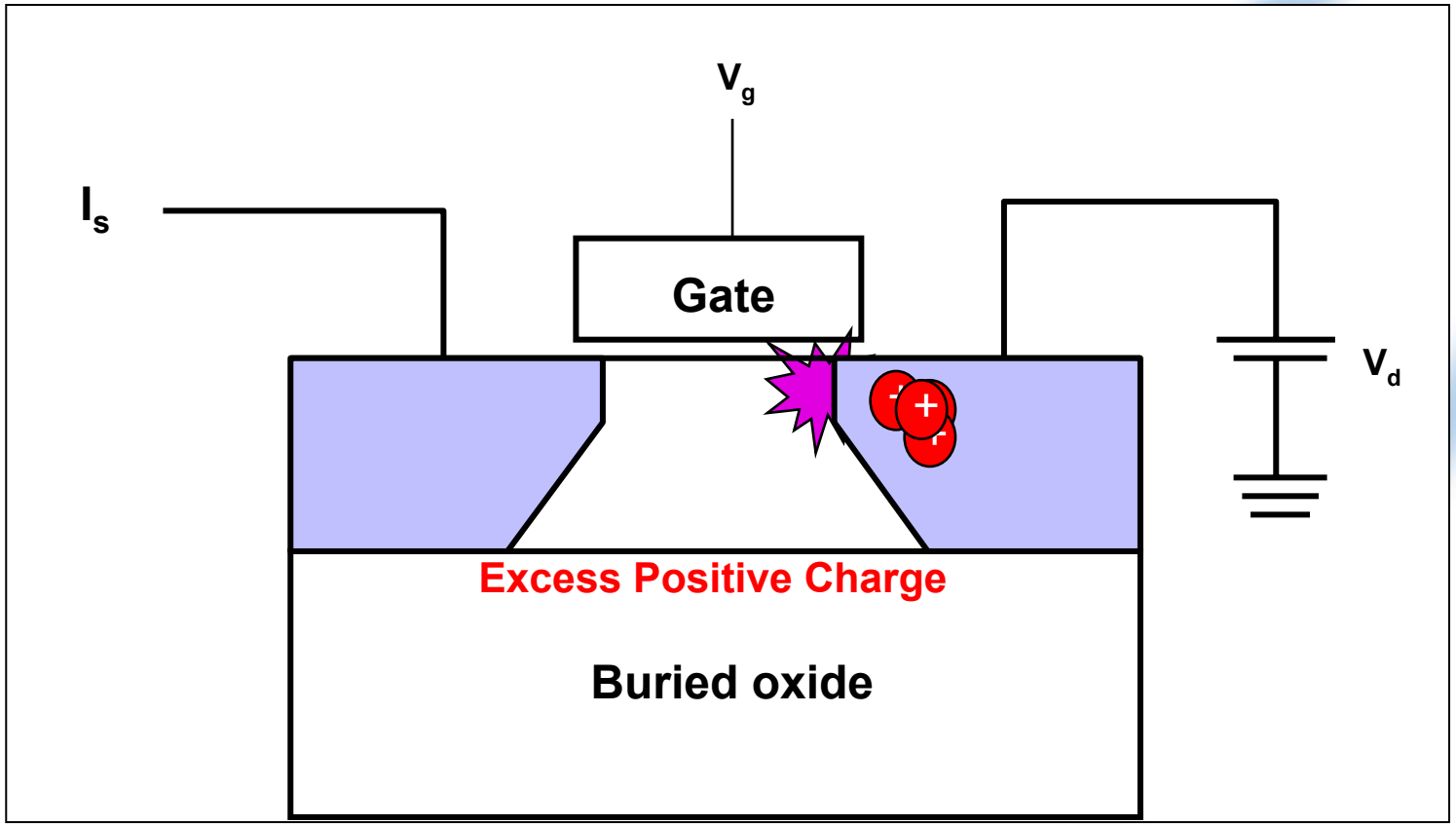


Logic "0"

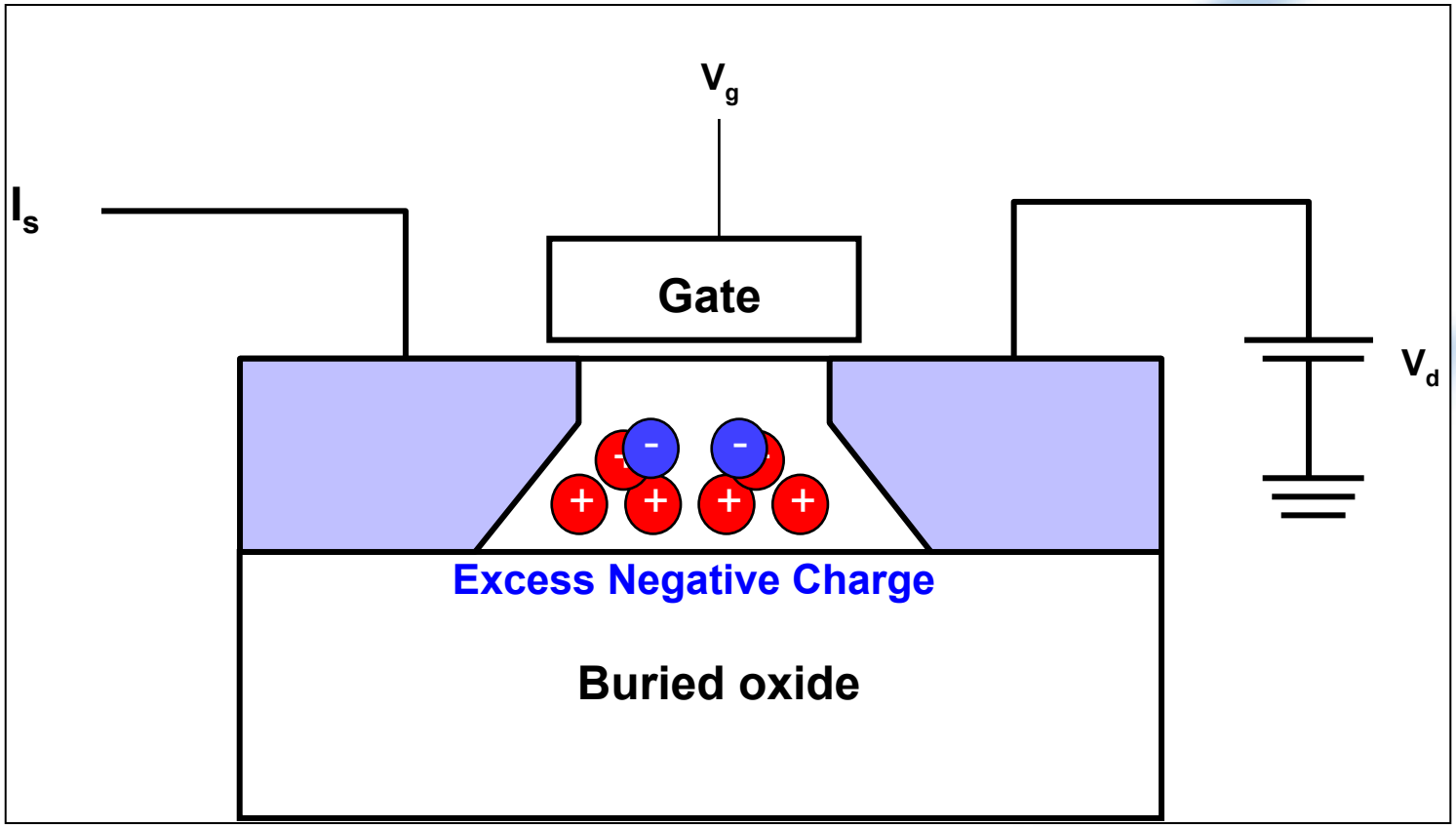




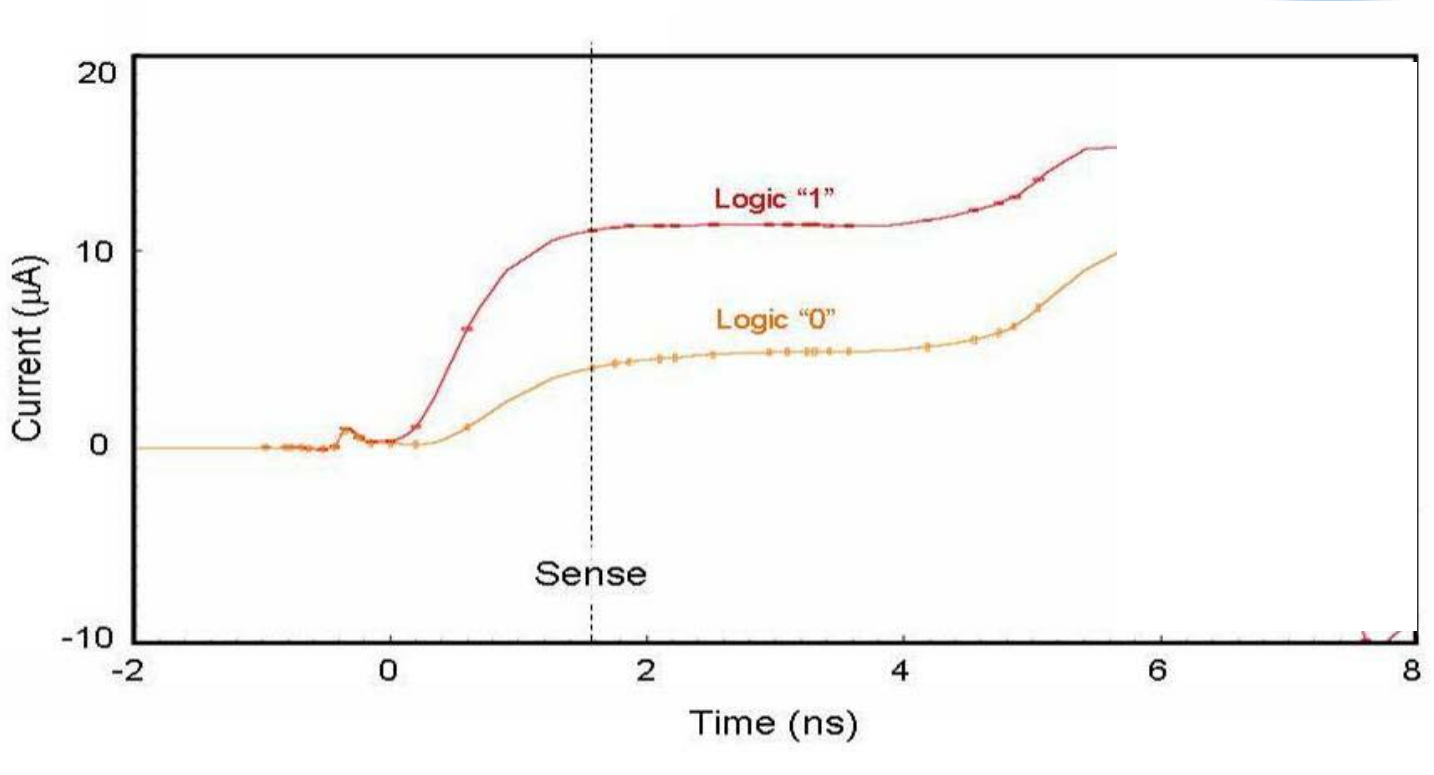
Z-RAM Device Overview:
Writing "1" by Impact Ionization



Z-RAM Device Overview: Writing "0" by Hole Removal



Z-RAM Device Overview:
Bit cell Read Current Waveforms

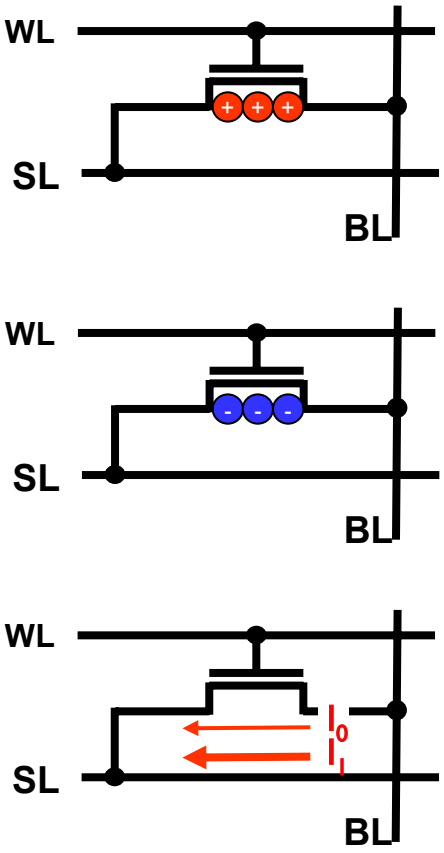


Z-RAM Device Overview:

Z-RAM vs. eDRAM

Innovative Silicon inc

Z-RAM

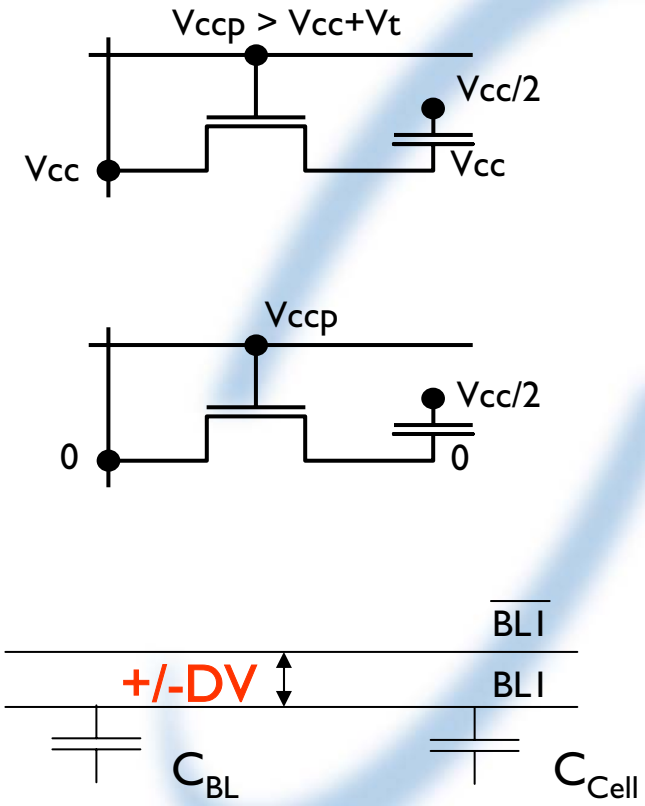


Store "1"

Store "0"

Read

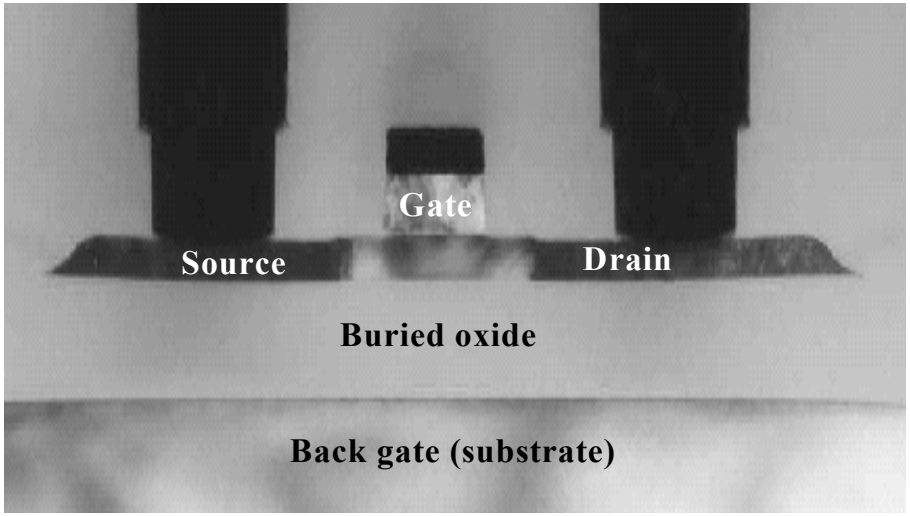
eDRAM



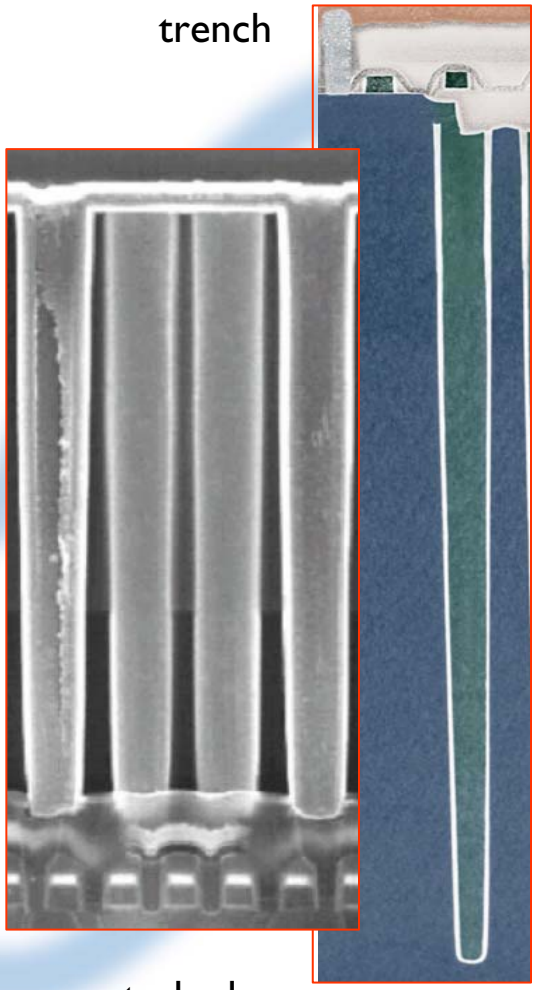
Z-RAM Memory Cell Operation

- Very Similar to eDRAM but with benefits of
  - Simple Process Integration
  - Robust Read method
  - Scalability
    - No Capacitor
    - No New materials
    - Litho Friendly

SOI



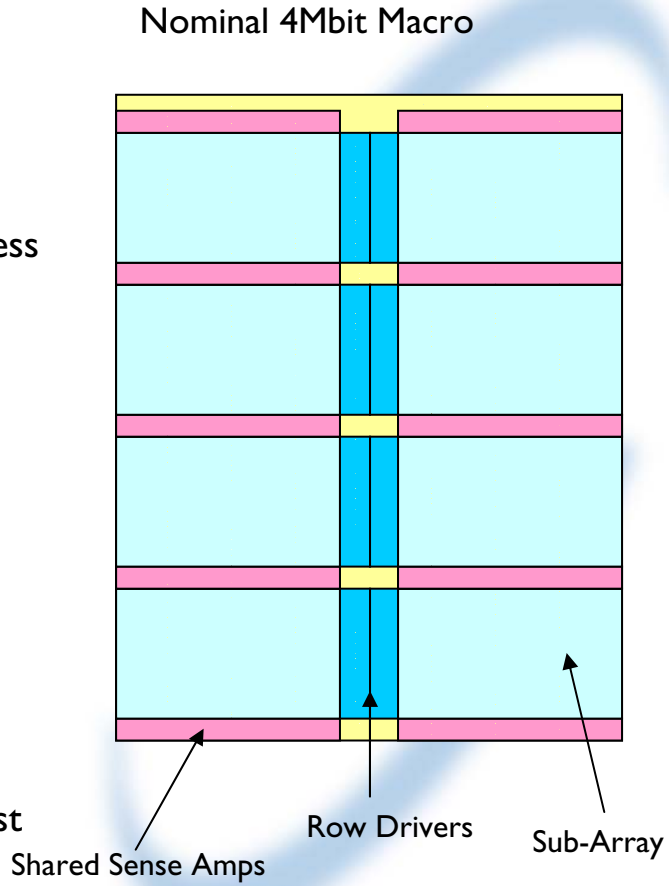
trench



stacked

Z-RAM Memory Macro Architecture Features

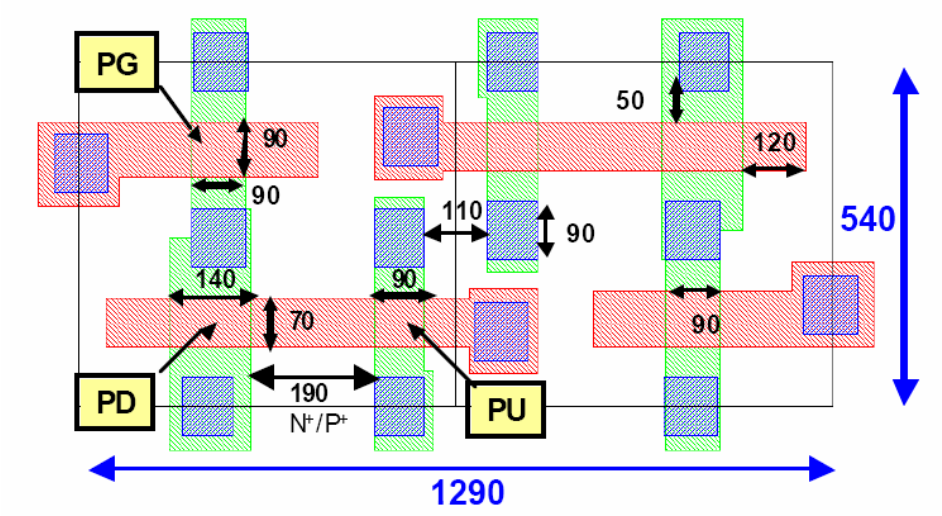
- Small Cell Size:
  - Enables Dense Sub-Arrays
  - Minimal RC loss
- Current Sensing
  - Allows long Bitlines with minimal impact on Read Access Time
  - Minimal Bitline swing during read
- Bitline Pitch limited by Metal minimum Spacing
  - Allows many bitlines along wordline
  - Large Page Size Possible from One Sub-array
- Global I/O
  - Best at ≥ 4x Sense Amp Pitch
  - SA-I/O MUX determines power overhead
- Burst operation Improves Efficiency
  - Faster, low power column (page) cycles
  - Refresh and Write-Back in Parallel with Data Out Burst



Z-RAM Memory Macro Architecture Features

Innovative Silicon inc

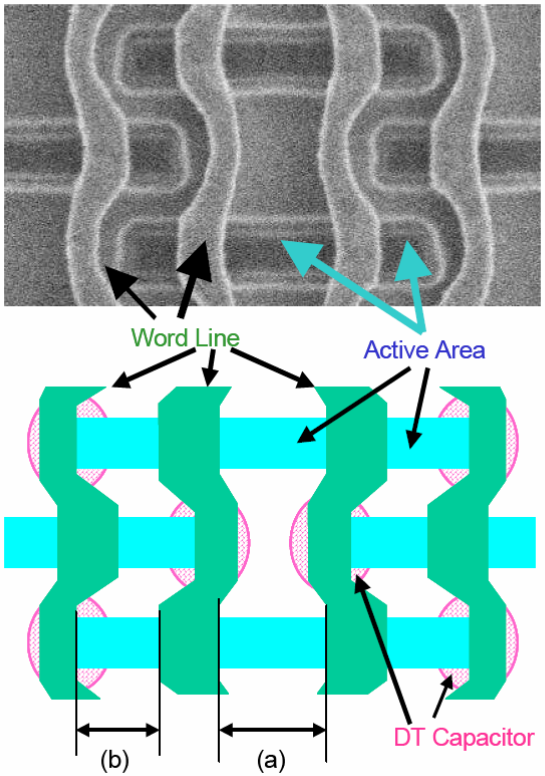
Example SRAM Cell, 65nm



A Functional 0.69μm Embedded 6T-SRAM bit cell for 65nm CMOS platform, Arnoud etal, VLSI 2003

Example DRAM Cell

(A) Top Down SEM of DRAM Cell



(B) DRAM Cell Layout

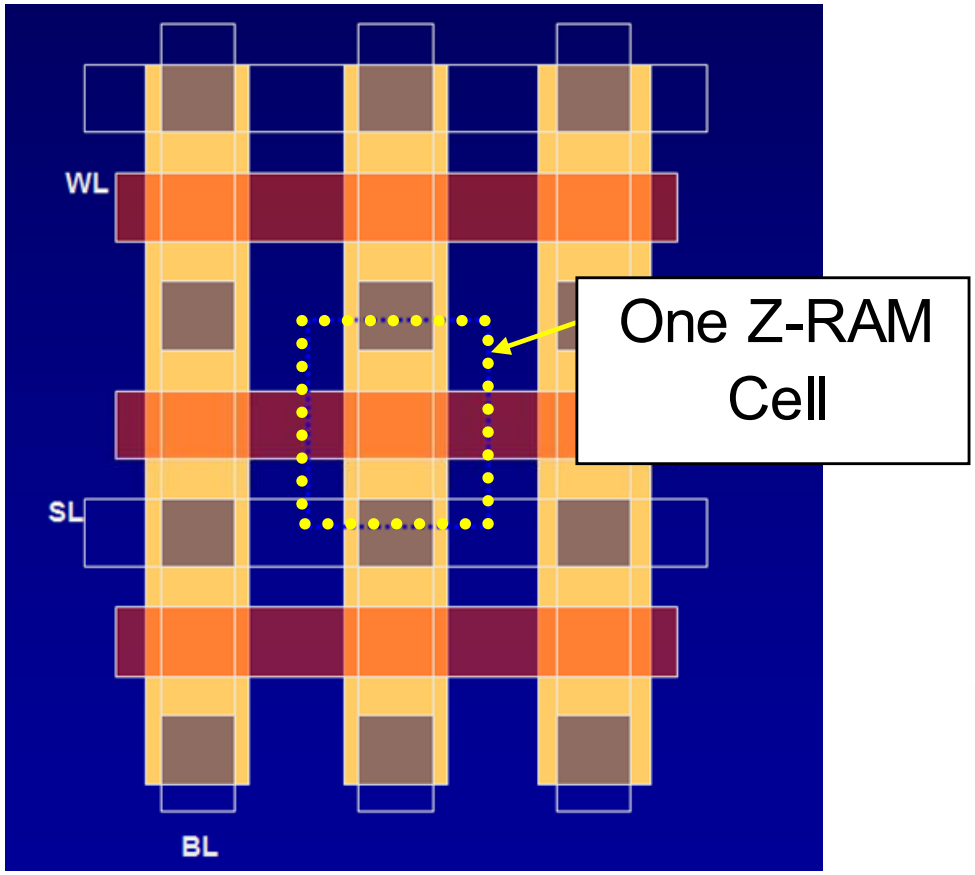
Fully Compatible Integration of High Density Embedded DRAM with 65nm CMOS Technology (CMOS5)

Y. Matsubara, etal., 2003, IEEE

Z-RAM Memory Macro Architecture Features

Innovative Silicon inc

Z-RAM Cell: Litho Friendly



Challenges

- Small Cell Size
  - Difficult to maintain array efficiency
  - Difficult to make Sense Amplifiers as fast as embedded SRAM equivalent
  - Random Access time slower than eSRAM
  - Can higher density and shorter routing compensate?
- More overhead with Restore & Refresh
  - SRAM does not require Restore/Refresh
  - Can Pipeline Operation or other architecture changes Address this?
- Power
  - Active Power and Standby Power often go up with improved array efficiency
  - How can the Z-RAM be optimized?

Z-RAM Performance Factors

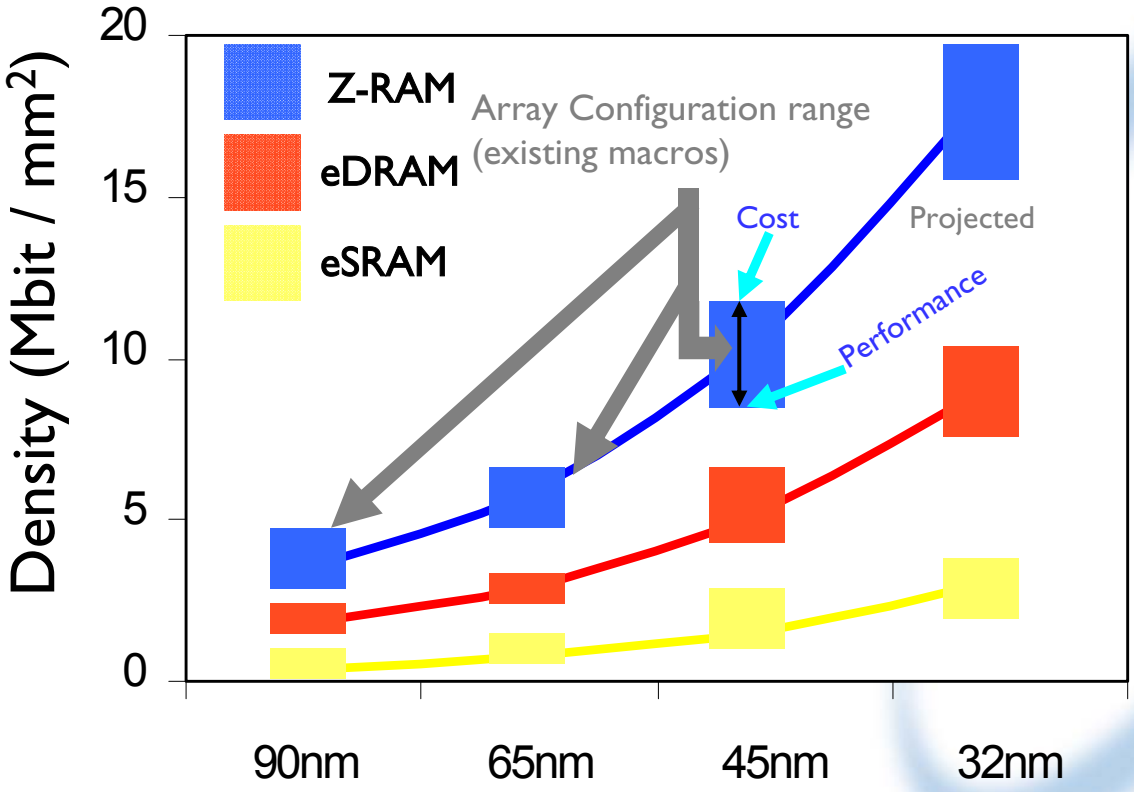
Innovative Silicon inc

- Memory Density vs. Speed
- Pipeline Operation
- Routing Impact
- Active Power
- Standby Power

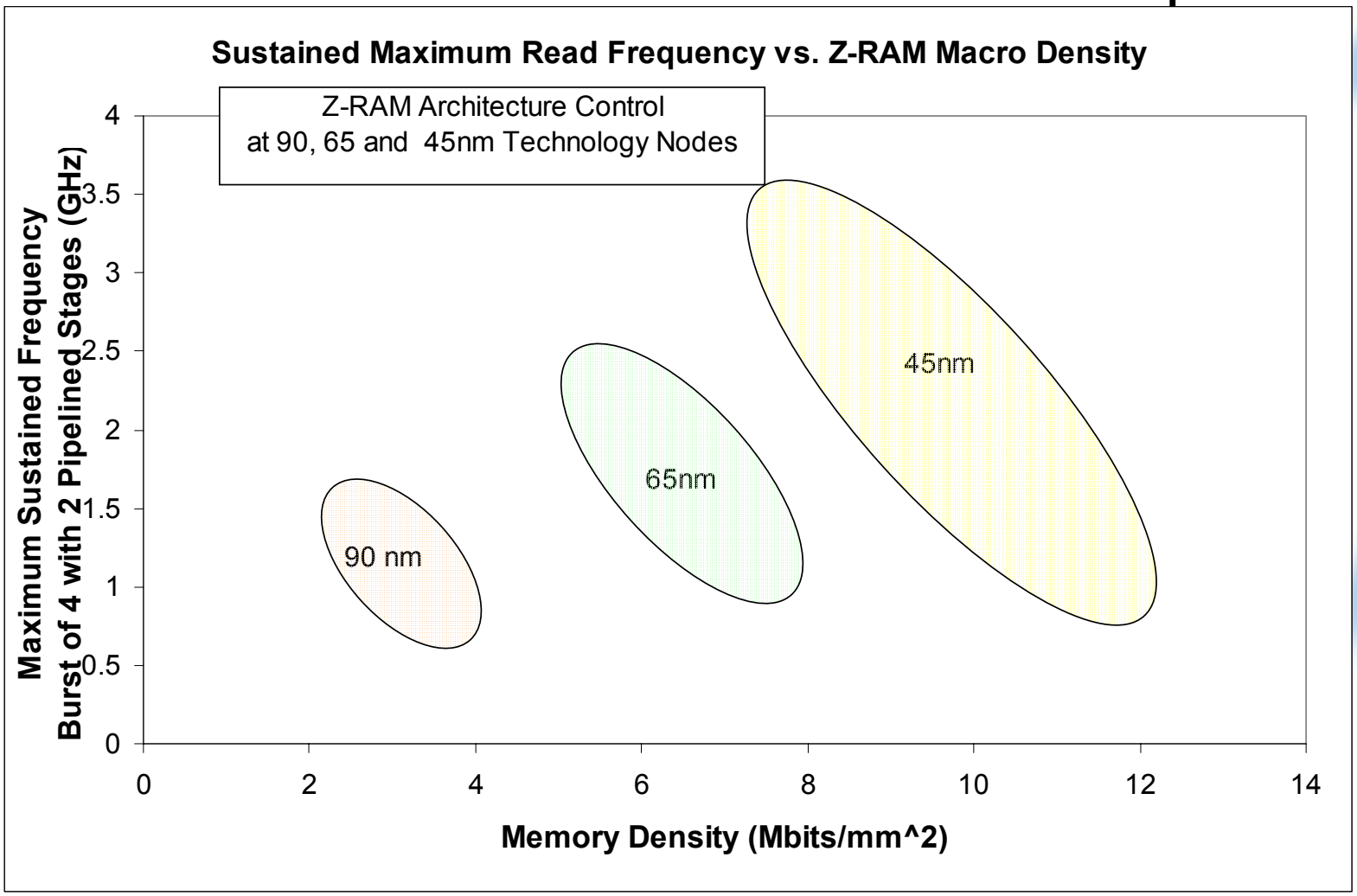
Z-RAM Performance Factors

Innovative Silicon inc

- Memory Density Comparison



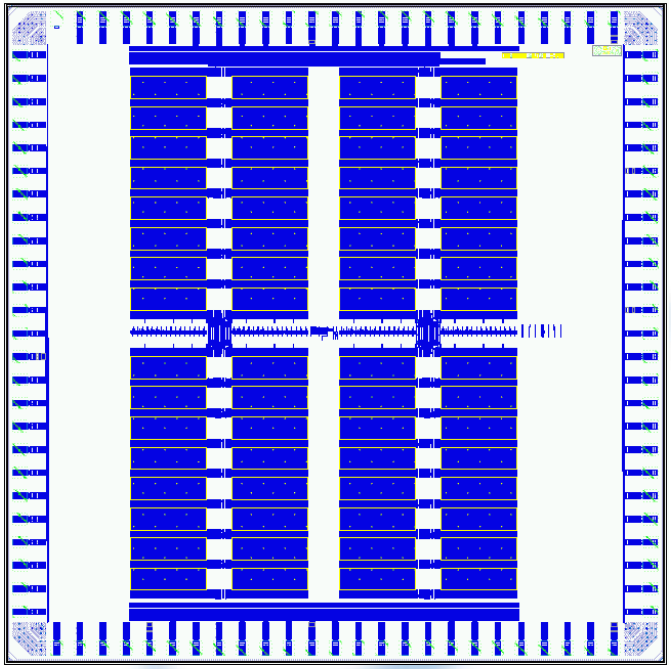
Z-RAM Performance Factors



Z-RAM Performance Factors
Memory Density vs. Speed

Innovative Silicon inc

- Access Time Better than eDRAM and Approaches SRAM
  - Benefit of underlying SOI technology
  - Low Voltage Wordline Swing
  - Minimal Bitline Voltage Swing
  - Short Interconnect
- Configure to reduce Access Time?
  - Obtain Sub-Array address early
    - Precharge Sub-Array Bitlines
    - Build with Smaller Sub Arrays

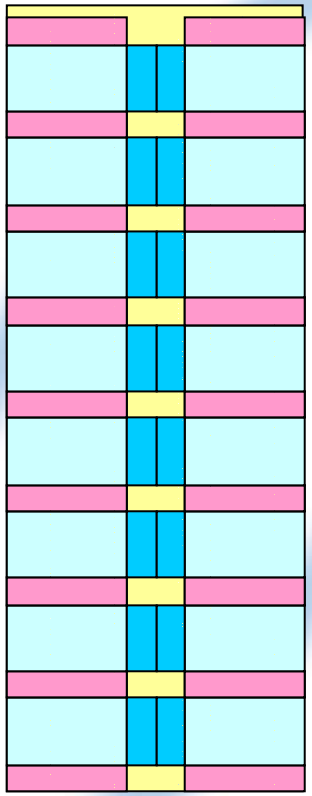


Example 16meg Z-RAM Macro

Z-RAM Performance Factors
Memory Density vs. Speed

- Limitations
  - Sense Amp Pitch vs. Speed
  - Array Efficiency
  - Total Cycle Time
  - Write-Back & Precharge
- Relative to SRAM:
  - Density: 4 - 5 x Higher
  - Access Time: 20 -30% Longer
- Solution Depends on Application
- System Performance vs. Memory Performance depends on Application

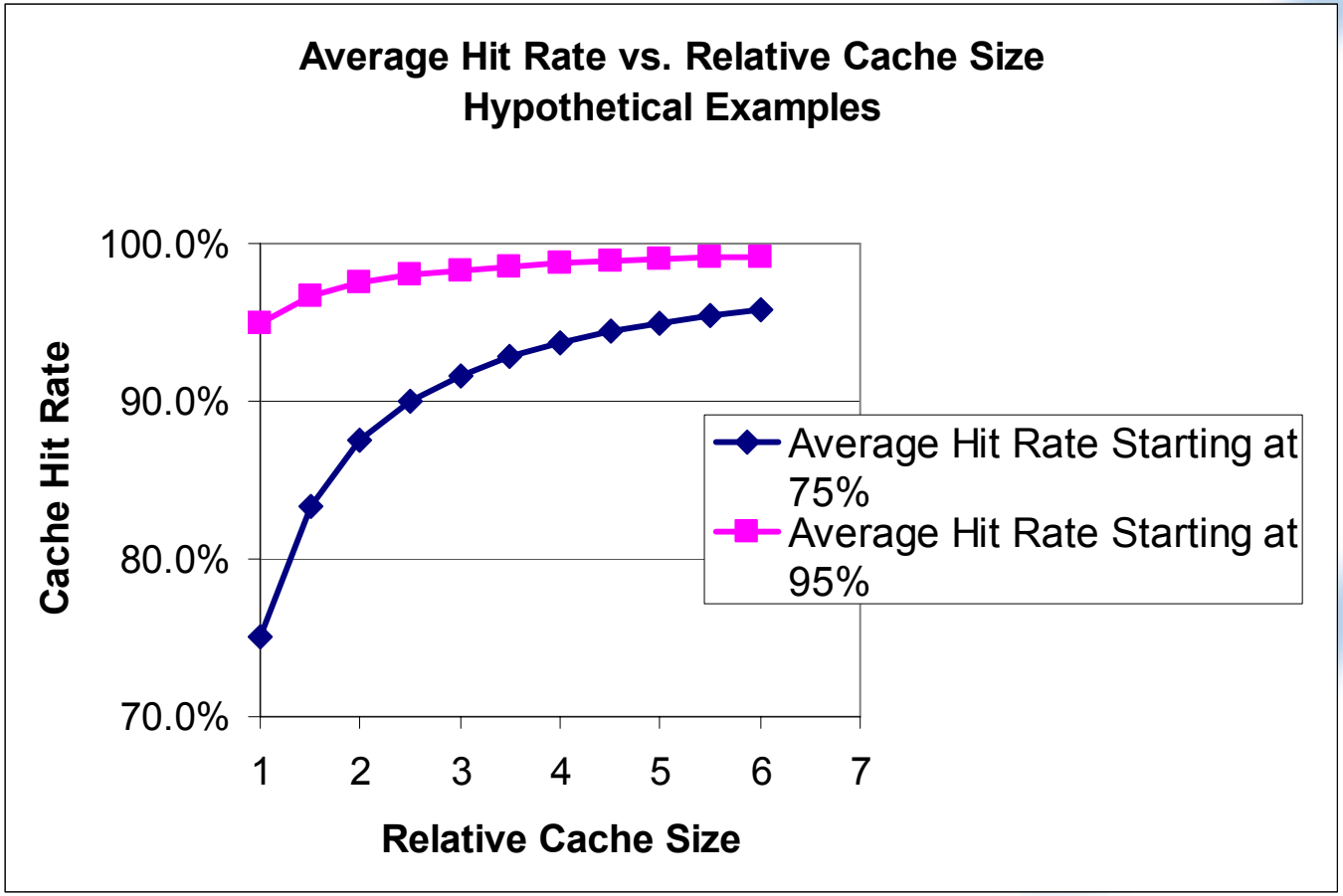
High Speed 2Mbit Macro



Z-RAM Performance Factors
Memory Density vs. Speed

Innovative Silicon inc

- Step 1: How does your Cache Size Impact Hit Rate?



Z-RAM Performance Factors
Memory Density vs. Speed

Innovative Silicon inc

- Step 2: What is the penalty of a "Miss"?
  - Typical Rule of Thumb: 10x or Greater Access Time Penalty
  - Depends on Application Pipeline/Architecture
    - Penalty could be greater with Deeper Pipelines

Z-RAM Performance Factors
Memory Density vs. Speed

Innovative Silicon inc

- Step 3: What is the Average Latency?
  - Average Latency

    =HR * (Ta1) + (1-HR)*(Ta2)

    where HR = Hit Rate, Ta1= Hit access Time, Ta2= Miss access Time

  Assume Ta2= Ta1 * PF

    where PF = Miss Access Time Penalty Factor

    =HR* (Ta1) + (1-HR)*(Ta1)*(PF)

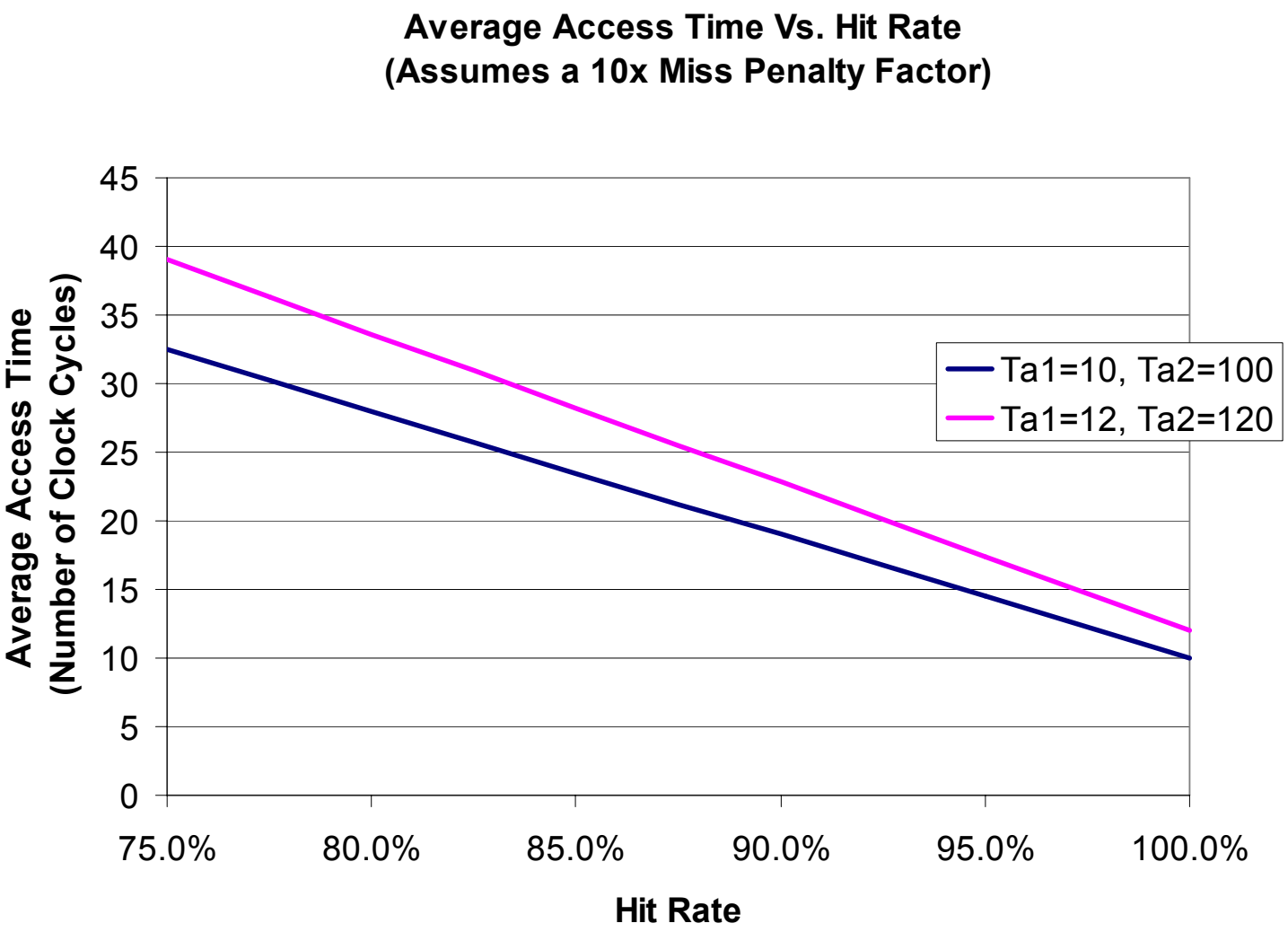
    =(Ta1)* (HR + (1-HR)* (PF))

    =(Ta1)*(PF - (PF-1)*HR

  If PF = 10

    = Ta1*(10 – 9*HR)

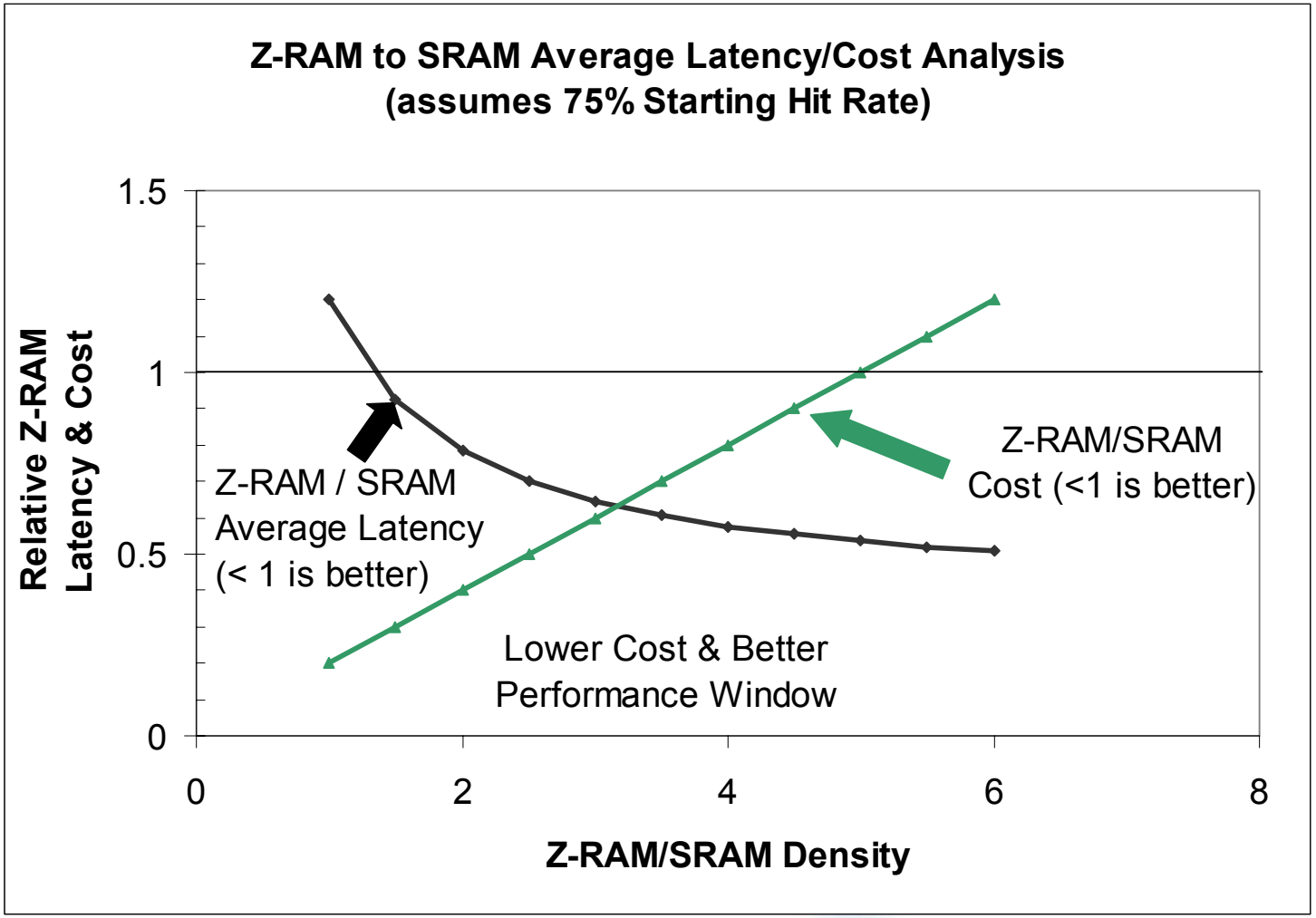
Z-RAM Performance Factors
Memory Density vs. Speed



Z-RAM Performance Factors
Memory Density vs. Speed

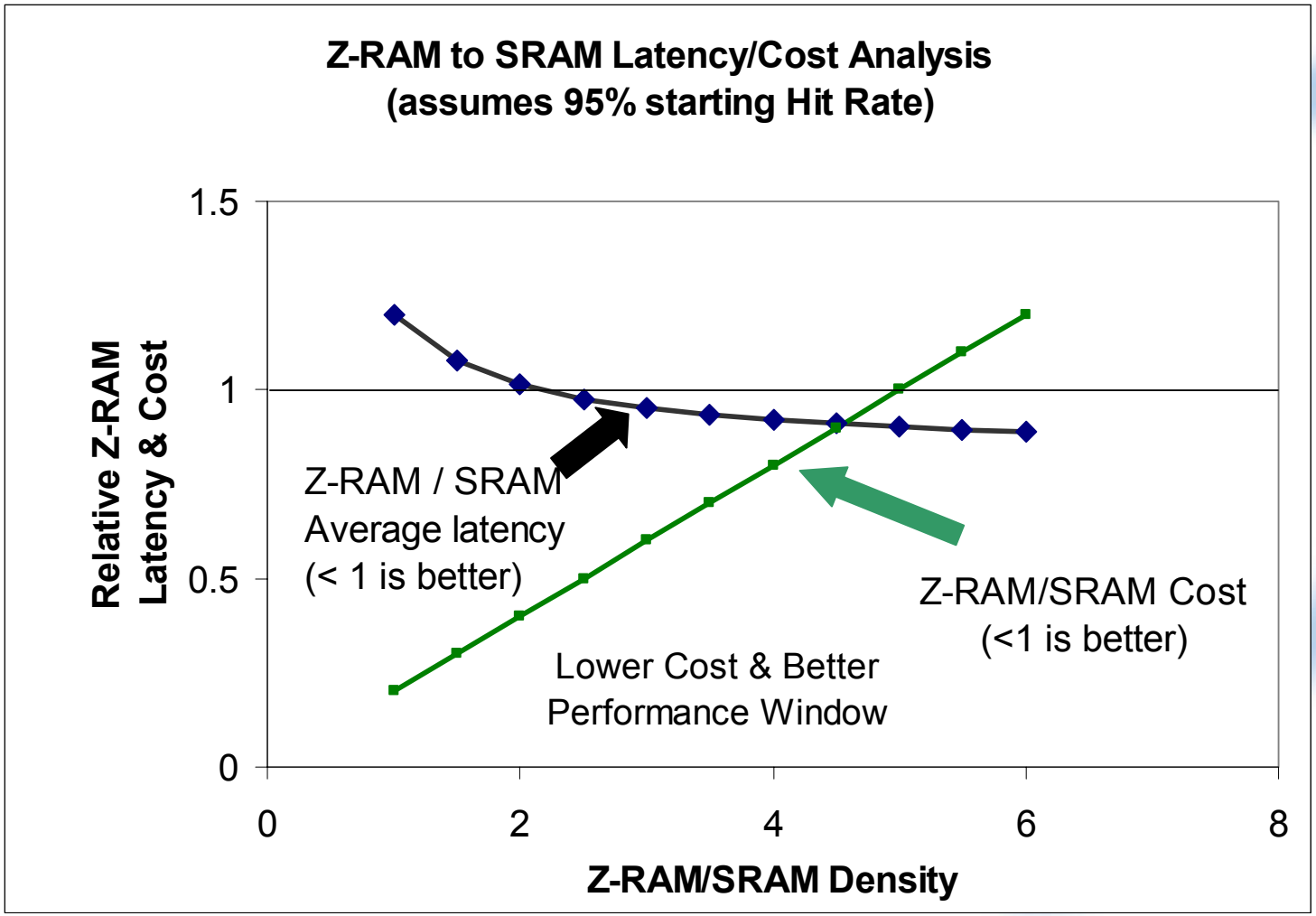
Innovative Silicon inc

- Step 4: Compare & Analyze



Larger yet 'slower' Cache can yield Lower Cost and Better Performance

Z-RAM Performance Factors
Memory Density vs. Speed



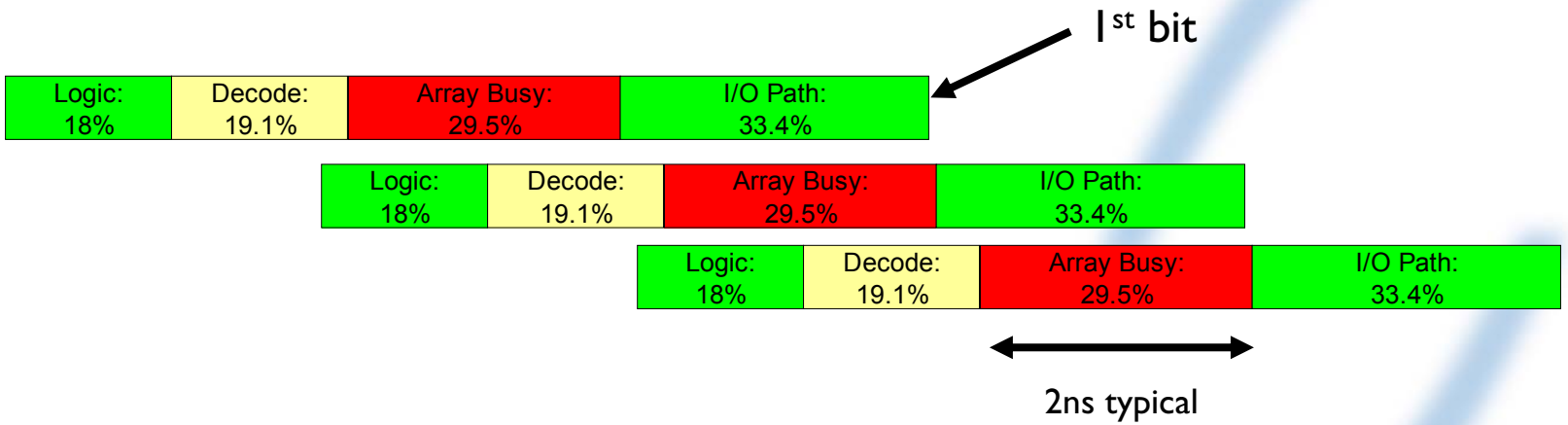
Z-RAM Performance Factors

Innovative Silicon inc

- Memory Density vs. Speed
- Pipeline Operation
- Routing Impact
- Active Power
- Standby Power

Z-RAM Performance Factors
Pipeline Operation

- Array 'busy' time is typically 30% or less of the total access time
- This provides opportunity for pipelined operation (similar to eDRAM & eSRAM)
- Pipeline Operation Increases Maximum frequency by 3x
- Typical Internal Timing Breakdown with 256K Sub-arrays (65nm):



•Burst of 4 yields 1.7Ghz sustained Output; limited by I/O path

•Logic and Array timing can be made faster (larger drivers, smaller sub-arrays)

Z-RAM Performance Factors
Pipeline Operation

Innovative Silicon inc

- Other Options:
  - Multibank and or Burst Read Operation
  - Refresh During Burst
  - Write-Back During Burst
  - Controller: Is Write-Back Required?
  - Will the data be used again?
  - If not, then Write-back can be skipped

Z-RAM Performance Factors

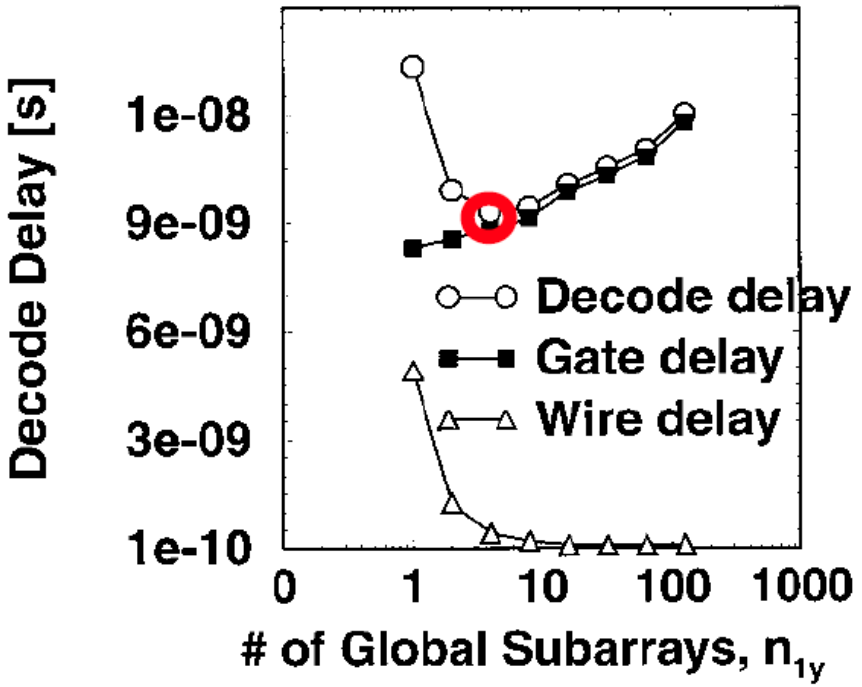
Innovative Silicon inc

- Memory Density vs. Speed
- Pipeline Operation
- Routing Impact
- Active Power
- Standby Power

Z-RAM Performance Factors
Routing Impact

Innovative Silicon inc

- Key Parameters is total turn-around time for a memory access
  - Direct memory access time: 25% to 50% of Total Time
  - SRAM –Many Steps Required to Reduce Wire Delay
    - More Banks
    - Local & Global Sensing
- How much Does a Smaller Footprint Help?
  - Wire Delay across Macro can still be significant
  - Z-RAM: RC delays ~ 50% of SRAM
  - Can Translate into 5-15% Access time Savings



Reference: Interconnect Centric Array Architectures for Minimum SRAM Access Time, Bhavnagarwala etal., ICCD'01

Z-RAM Performance Factors

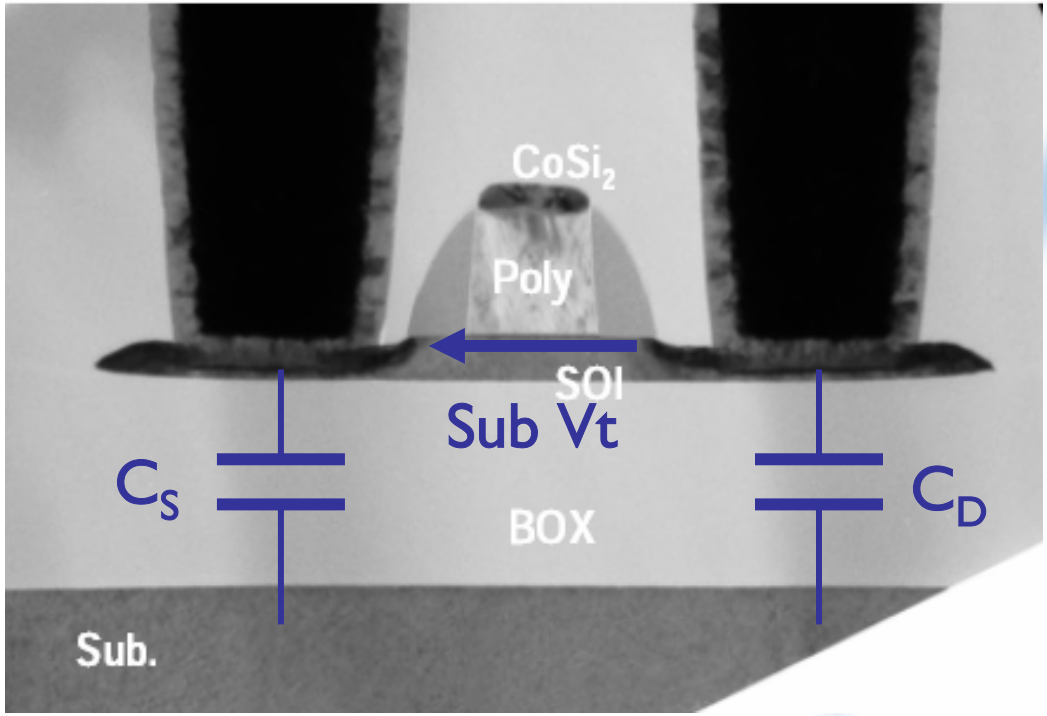
Innovative Silicon inc

- Memory Density vs. Speed
- Pipeline Operation
- Routing Impact
- Active Power
- Standby Power

Z-RAM Performance Factors
Active Power

Innovative Silicon inc

- Benefit of SOI Process



Low parasitic S/D cap., near ideal sub Vt slope
Operating power 3 times smaller vs. Bulk

Ref.: Y. Nagatomo, IEEE International SOI Conf., 2005

Z-RAM Performance Factors
Active Power

Innovative Silicon inc

- Power Can Be Modulated by Architecture
- Can Reduce Active Power with:
  - Large Line Size
  - Longer Burst
  - Shorter Bitlines
  - Z-RAM Architecture allows large Line Size per Sub Array
  - Cell Pitch limited only by Bitline to Bitline Pitch

Z-RAM Performance Factors

Innovative Silicon inc

- Memory Density vs. Speed
- Pipeline Operation
- Routing Impact
- Active Power
- Standby Power

Z-RAM Performance Factors
Standby Power

Innovative Silicon inc

- Standby Current composed of Refresh Current and Peripheral Leakage Current
- Array Leakage is minimal
- In Standby:
  - Source= 0 Volts
  - Drain = 0 Volts
  - Gate = 0 Volts
  - 65nm SRAM Standby Power ~ 93mW/Mbit
    - Array Remains Powered Up
    - (*16 MB L3 Cache on Intel Xeon processor, VLSI Conference 2006*)
  - At 65nm, total Z-RAM Standby Power is 50-90% Lower than SRAM Standby Power

Summary

Innovative Silicon inc

- Z-RAM Embedded Memory is an exciting Technology
  - Compatible with standard SOI logic or memory process - No new materials
  - Small Cell Size/Low Cost
  - Scalable
  - Works for N and PMOS
  - Good SER Performance
  - Low Power
  - **Application Configurable**